

### PRÁCTICA DIRIGIDA 4

#### Pregunta 1 – Tobit<sup>1</sup>

Si deseamos estimar un modelo que explique las horas trabajadas por una mujer casada, ¿qué variables explicativas incluiríamos? La literatura empírica ha demostrado que algunos de los factores que tienden a atraer a una mujer a la fuerza de trabajo son su educación y su experiencia previa en el mercado laboral. Por otro lado, los factores que podrían reducir su incentivo para trabajar son su edad y la presencia de niños pequeños en el hogar. Por lo tanto, podríamos proponer el siguiente modelo de regresión:

$$Hours_i = \beta_0 + \beta_1 Educ_i + \beta_2 Exper_i + \beta_3 Age_i + \beta_4 KIDSL6_i + \varepsilon_i$$

donde la variable observada HOURS es un número positivo de horas trabajadas, o cero, para las mujeres que no ingresaron a la fuerza laboral. Por otro lado, KIDSL6 es la cantidad de niños menores de 6 años en el hogar.

Para esta pregunta se le pide usar los datos del estudio de Thomas Mroz (1987)<sup>2</sup> sobre la participación de las mujeres casadas en la fuerza laboral. Los datos están en el archivo *mroz.dta*, la muestra consiste de 753 observaciones sobre mujeres casadas entre 30 y 60 años en 1975. De este grupo, 325 no trabajaron fuera del hogar y, por lo tanto, no reportaron horas trabajadas ni salarios.

Se pide realizar lo siguiente:

- 1) Describa y examine los estadísticos más importantes de las variables.

---

<sup>1</sup> El desarrollo de esta pregunta se encuentra en el do file "tobit\_pregunta1"

<sup>2</sup> Mroz, T.A. (1987): "The Sensitivity of an Empirical Model of Married Women's Hours of Work to Economic and Statistical Assumptions", *Econometrica*, 55, 765-799.

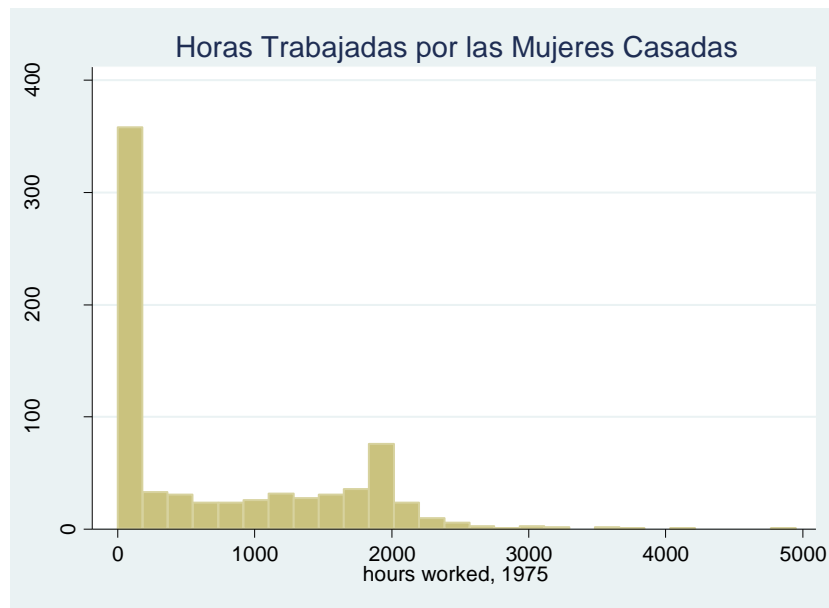
```
. describe hours educ exper age kidslt6
```

variable name	storage type	display format	value label	variable label
hours	int	%9.0g		hours worked, 1975
educ	byte	%9.0g		years of schooling
exper	byte	%9.0g		actual labor mkt exper
age	byte	%9.0g		woman's age in yrs
kidslt6	byte	%9.0g		# kids < 6 years

```
. sum hours educ exper age kidslt6
```

Variable	Obs	Mean	Std. Dev.	Min	Max
hours	753	740.5764	871.3142	0	4950
educ	753	12.28685	2.280246	5	17
exper	753	10.63081	8.06913	0	45
age	753	42.53785	8.072574	30	60
kidslt6	753	.2377158	.523959	0	3

2) Analice la variable *hours* usando un histograma. ¿Qué características presenta?



El gráfico anterior muestra el histograma de horas trabajadas. El histograma muestra que hay una gran proporción de mujeres casadas que no participaron en el mercado laboral en 1975, por ello, reportaron cero horas de trabajo. Este es un ejemplo de datos censurados, ya que una fracción significativa de las observaciones sobre la variable dependiente toma un valor límite, que en este caso es cero para aquellas mujeres que no trabajaron.

- 3) Repita el mismo análisis de la pregunta 1 considerando solo aquellas personas que reportaron una cantidad positiva de horas trabajadas ( $hours > 0$ ), y compárelas con los estadísticos descriptivos de las mujeres que no se encuentran en la fuerza laboral ( $hours = 0$ ).

```
. sum hours educ exper age kidslt6 if (hours>0)
```

Variable	Obs	Mean	Std. Dev.	Min	Max
hours	428	1302.93	776.2744	12	4950
educ	428	12.65888	2.285376	5	17
exper	428	13.03738	8.055923	0	38
age	428	41.97196	7.721084	30	60
kidslt6	428	.1401869	.3919231	0	2

```
. sum hours educ exper age kidslt6 if (hours==0)
```

Variable	Obs	Mean	Std. Dev.	Min	Max
hours	325	0	0	0	0
educ	325	11.79692	2.181995	5	17
exper	325	7.461538	6.918567	0	45
age	325	43.28308	8.467796	30	60
kidslt6	325	.3661538	.6368995	0	3

Podemos observar que las mujeres que no trabajan tienen en promedio una educación y experiencia menor, son un poco mayores y tienen más hijos pequeños menores de 6 años, respecto de las mujeres que participan en el mercado laboral.

- 4) Realice una regresión de mínimos cuadrados usando primero todas las observaciones de la muestra, y luego realice una nueva estimación usando solo a la sub muestra de mujeres que reportaron un cantidad de horas trabajadas positivas. ¿Qué problema potencial enfrentan los coeficientes estimados?

Regresión por MCO usando la muestra completa:

```
. reg hours educ exper age kidslt6
```

Source	SS	df	MS	Number of obs = 753		
Model	146771295	4	36692823.7	F( 4, 748) = 64.71		
Residual	424138429	748	567029.985	Prob > F = 0.0000		
				R-squared = 0.2571		
				Adj R-squared = 0.2531		
Total	570909724	752	759188.463	Root MSE = 753.01		

hours	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
educ	27.08568	12.23989	2.21	0.027	3.057054	51.1143
exper	48.03981	3.641804	13.19	0.000	40.89044	55.18919
age	-31.30782	3.96099	-7.90	0.000	-39.0838	-23.53184
kidslt6	-447.8547	58.41252	-7.67	0.000	-562.5267	-333.1827
_cons	1335.306	235.6487	5.67	0.000	872.6945	1797.918

Regresión por MCO usando la muestra no censurada (mujeres con *hours* > 0):

```
. reg hours educ exper age kidslt6 if (hours>0)
```

Source	SS	df	MS	Number of obs = 428		
Model	32193987.4	4	8048496.86	F( 4, 423) = 15.12		
Residual	225117032	423	532191.566	Prob > F = 0.0000		
				R-squared = 0.1251		
				Adj R-squared = 0.1168		
Total	257311020	427	602601.92	Root MSE = 729.51		

hours	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
educ	-16.46211	15.58083	-1.06	0.291	-47.0876	14.16339
exper	33.93637	5.009185	6.77	0.000	24.09038	43.78237
age	-17.10821	5.457674	-3.13	0.002	-27.83575	-6.380677
kidslt6	-305.309	96.44904	-3.17	0.002	-494.8881	-115.7299
_cons	1829.746	292.5356	6.25	0.000	1254.741	2404.75

Los estimadores de mínimos cuadrados no son confiables ya que el estimador de MCO es sesgado como inconsistente.

- 5) Ahora utilice un Tobit para estimar el mismo modelo. Compare los resultados con los de MCO.

Utilizando el modelo Tobit que primero estima la probabilidad de que la cantidad de horas trabajadas sea igual a cero (*hours* = 0) versus (*hours* > 0), y para aquellas mujeres con una cantidad de horas mayor a cero estima los parámetros de interés, se obtienen los siguientes resultados.

```
. tobit hours educ exper age kidslt6, ll(0)
```

```
Tobit regression                                Number of obs   =          753
                                                LR chi2(4)      =        255.50
                                                Prob > chi2     =         0.0000
Log likelihood = -3827.1433                    Pseudo R2      =         0.0323
```

hours	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
educ	73.29099	20.47458	3.58	0.000	33.09659	113.4854
exper	80.53527	6.287805	12.81	0.000	68.19145	92.87909
age	-60.7678	6.88819	-8.82	0.000	-74.29025	-47.24534
kidslt6	-918.9181	111.6606	-8.23	0.000	-1138.123	-699.713
_cons	1349.876	386.2989	3.49	0.001	591.5188	2108.234
/sigma	1133.697	42.06234			1051.123	1216.271

```
Obs. summary:      325 left-censored observations at hours<=0
                   428 uncensored observations
                   0 right-censored observations
```

Los coeficientes estimados mediante el modelo Tobit tiene los signos esperados y todos son estadísticamente significativos. Cabe señalar que la interpretación de los coeficientes, al igual que en las estimaciones de modelos Probit, se debe hacer con los efectos marginales. Sin embargo, debemos tener cuidado ya que en el modelo Tobit existen distintos tipos de efectos marginales.

- 6) Calcule los efectos marginales de cada variable explicativa sobre la probabilidad de que una mujer tenga una cantidad positiva de horas trabajadas.

```
. mfx compute, predict(pr(0,.))
```

```
Marginal effects after tobit
      y  = Pr(hours>0) (predict, pr(0,.))
      = .60542747
```

variable	dy/dx	Std. Err.	z	P> z	[ 95% C.I. ]		X
educ	.0248848	.00695	3.58	0.000	.011271	.038499	12.2869
exper	.0273445	.00219	12.51	0.000	.023061	.031628	10.6308
age	-.0206328	.00236	-8.75	0.000	-.025254	-.016012	42.5378
kidslt6	-.3120045	.03823	-8.16	0.000	-.386937	-.237071	.237716

Vemos que un año adicional de educación aumenta la probabilidad de que la cantidad de horas trabajadas sea mayor a cero ( $hours > 0$ ) en un 2.48%, asimismo, un año adicional de experiencia aumenta la probabilidad de que las mujeres trabajen en un 2.73%. Por otro lado, un año adicional de edad disminuye la probabilidad de que las mujeres trabajen en un 2%, y por último tener un hijo adicional menor a 6 años disminuye la probabilidad de que la mujer se encuentre trabajando en un 31%.

- 7) Calcule el efecto marginal de cada variable explicativa en el esperado de las horas trabajadas condicional a que la mujer se encuentra trabajando en el mercado laboral  $E(Hours|Hours > 0)$

```
. mfx compute, predict(e(0,.))
```

Marginal effects after tobit

```
y = E(hours|hours>0) (predict, e(0,.))
= 1023.9761
```

variable	dy/dx	Std. Err.	z	P> z	[ 95% C.I. ]		x
educ	31.2026	8.67013	3.60	0.000	14.2095	48.1957	12.2869
exper	34.28675	2.63567	13.01	0.000	29.1209	39.4526	10.6308
age	-25.87103	2.88484	-8.97	0.000	-31.5252	-20.2168	42.5378
kidslt6	-391.2163	46.689	-8.38	0.000	-482.725	-299.708	.237716

## Pregunta 2 – Selección Muestral

“La selección muestral aparece cuando la inclusión de una unidad económica en la muestra depende de decisiones previamente tomadas por dicha unidad, por lo que la muestra no debe considerarse aleatoria. Este es el caso de un análisis del salario percibido por las mujeres trabajadoras, pues sólo podrán informar acerca de su salario aquellas mujeres que hayan decidido trabajar por percibir un salario superior a su salario de reserva.” (Novales, 1993)

Para esta pregunta se le pide usar nuevamente los datos del estudio de Thomas Mroz (1987) sobre la participación de las mujeres casadas en la fuerza laboral. Los datos están en el archivo *mroz.dta*, la muestra consiste de 753 observaciones sobre mujeres casadas, de las cuales 325 no trabajaron y, por lo tanto, no reportaron horas trabajadas ni salarios.

Estamos interesados en estimar los retornos de la educación en los salarios de las mujeres tomando en cuenta el sesgo de selección o selección muestral que puede existir en los datos. Para ello se le pide desarrollar las siguientes preguntas:

**1) Estime por Mínimos Cuadrados un modelo para el logaritmo del salario  $\ln(Wage)$  usando como variables explicativas la educación de la mujer (EDUC) y los años de experiencia laboral (EXPER), considerando solo a las 428 mujeres que trabajaron.**

```
. regress lwage educ exper if (hours>0)
```

Source	SS	df	MS	Number of obs = 428		
Model	33.1324581	2	16.566229	F( 2, 425) = 37.02		
Residual	190.194983	425	.447517607	Prob > F = 0.0000		
				R-squared = 0.1484		
				Adj R-squared = 0.1444		
Total	223.327441	427	.523015084	Root MSE = .66897		

lwage	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
educ	.1094888	.0141672	7.73	0.000	.0816423	.1373353
exper	.0156736	.0040191	3.90	0.000	.0077738	.0235733
_cons	-.4001744	.1903682	-2.10	0.036	-.7743548	-.0259939

Usando MCO obtenemos que el retorno estimado de la educación es 11%, asimismo los coeficientes estimados tanto de educación como de experiencia son estadísticamente significativos.

## 2) Estime el retorno de la educación para las mujeres casadas, usando la metodología de 2 etapas de Heckman para corregir el sesgo de selección, detallando cada uno de sus pasos.

### Primera Etapa (Estimación de la ecuación de selección usando un Probit)

El método de dos etapas de Heckman consiste en estimar en un primer paso un modelo tipo Probit para calcular la probabilidad de participar en el mercado laboral para una mujer, dadas ciertas variables de interés que determinen tal decisión. En particular, usaremos como variables explicativas de la probabilidad de que una mujer se encuentre trabajando la edad, los años de educación, una variable dummie que nos indique si tiene hijos o no, y la tasa de impuestos que le tocaría pagar si estuviese trabajando. Es importante señalar que las variables incluidas en la estimación de la “ecuación de selección” mediante un modelo Probit no pueden ser exactamente iguales a las variables que se incluirán en la estimación por MCO de la ecuación principal.

Primero, generamos una variable dicotómica que tome el valor de 1 si la mujer se encontraba trabajando, y 0 si no lo hizo. Esta variable dicotómica será usada como variable dependiente en la estimación del modelo Probit.

```
. generate dwork = 1

. replace dwork = 0 if lwage == .
(325 real changes made)
```

En seguida, procedemos a estimar la probabilidad de participar en el mercado laboral usando un modelo Probit:

```
. probit dwork age educ kids mtr
```

```
Iteration 0: log likelihood = -514.8732
Iteration 1: log likelihood = -494.21411
Iteration 2: log likelihood = -494.14614
Iteration 3: log likelihood = -494.14614
```

```
Probit regression                                Number of obs   =          753
                                                LR chi2(4)      =          41.45
                                                Prob > chi2     =          0.0000
Log likelihood = -494.14614                    Pseudo R2       =          0.0403
```

dwork	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
age	-.0206155	.0070447	-2.93	0.003	-.0344229	-.0068082
educ	.0837753	.023205	3.61	0.000	.0382943	.1292563
kids	-.3138848	.1237108	-2.54	0.011	-.5563535	-.0714162
mtr	-1.393853	.6165751	-2.26	0.024	-2.602318	-.1853877
_cons	1.192296	.7205439	1.65	0.098	-.2199444	2.604536

Como era de esperarse, los efectos de la edad, la presencia de hijos, y los impuestos más altos reducen significativamente la probabilidad de que una mujer participe en el mercado laboral. Por otro lado, la educación aumenta la probabilidad de que una mujer se una a la fuerza laboral.

Utilizando los coeficientes estimados mediante el modelo Probit, calcularemos el Inverso del Ratio de Mills, el cual captura la magnitud del sesgo de selección:

$$\tilde{\lambda} = IMR = \frac{\phi(1.1923 - 0.0206AGE + 0.0838EDUC - 0.3139KIDS - 1.3939MTR)}{\Phi(1.1923 - 0.0206AGE + 0.0838EDUC - 0.3139KIDS - 1.3939MTR)}$$

En Stata debemos usar primero el comando `predict` y luego generar una nueva variable para el inverso del Ratio de Mills. La función `normalden` es la pdf (Función de Densidad de Probabilidad) de la normal estándar, mientras que la función `normal` es la cdf (Función de Densidad Acumulada) de la normal estándar.

```
. predict w, xb
```

```
. gen InvMills = normalden(w)/normal(w)
```

### **Segunda Etapa (Estimación de la ecuación principal, incorporando como regresor adicional al Inverso del Ratio de Mills)**

Luego del cálculo del inverso del ratio de Mills, debemos estimar la ecuación principal por MCO, incluyendo al estimador del Inverso del Ratio de Mills como un regresor adicional, de tal manera que corrijamos el sesgo de selección. Como se mencionó anteriormente, en esta etapa no se pueden incorporar todas las variables incluidas en la estimación del modelo Probit, por ello, no incluimos la edad de las mujeres como variable explicativa en la ecuación de salarios.



```
. regress lwage educ exper InvMills
```

Source	SS	df	MS	Number of obs = 428		
Model	36.2307257	3	12.0769086	F( 3, 424) = 27.37		
Residual	187.096715	424	.441265838	Prob > F = 0.0000		
Total	223.327441	427	.523015084	R-squared = 0.1622		
				Adj R-squared = 0.1563		
				Root MSE = .66428		

lwage	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
educ	.0584579	.0238495	2.45	0.015	.01158	.1053358
exper	.0163202	.0039984	4.08	0.000	.0084612	.0241793
InvMills	-.8664387	.3269855	-2.65	0.008	-1.509153	-.2237242
_cons	.8105418	.4944723	1.64	0.102	-.1613804	1.782464

Es importante señalar los siguientes puntos: primero, el coeficiente estimado del inverso del Ratio de Mills (InvMills) es estadísticamente significativo, lo que implica que existe sesgo de selección en los resultados obtenidos por MCO estimados en la pregunta anterior. Segundo, el retorno estimado de la educación ha caído de 11% a cerca de 6%.

### 3) Estimar en Stata los retornos de la educación usando la metodología en 2 etapas de Heckman de manera directa.

El comando en Stata para realizar la metodología de 2 etapas de Heckman es el siguiente:

```
heckman lwage educ exper, select (dwork= age educ kids mtr) twostep
```

Este comando tiene dos partes: la primera corresponde a la ecuación de interés, y la segunda es la ecuación de selección, en donde se incorpora la variable dummie binaria creada, que toma el valor de 1 si hours>0 y toma el valor de 0 si la mujer no trabajó

```
. heckman lwage educ exper, select (dwork= age educ kids mtr) twostep
```

```
Heckman selection model -- two-step estimates      Number of obs      =          753
(regression model with sample selection)          Censored obs       =          325
                                                  Uncensored obs     =          428

                                                  Wald chi2(2)       =          19.53
                                                  Prob > chi2        =          0.0001
```

	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
lwage						
educ	.0584579	.0296354	1.97	0.049	.0003737	.1165422
exper	.0163202	.0042022	3.88	0.000	.0080842	.0245563
_cons	.8105417	.6107985	1.33	0.185	-.3866013	2.007685
dwork						
age	-.0206155	.0070447	-2.93	0.003	-.0344229	-.0068082
educ	.0837753	.023205	3.61	0.000	.0382943	.1292563
kids	-.3138848	.1237108	-2.54	0.011	-.5563535	-.0714162
mtr	-1.393853	.6165751	-2.26	0.024	-2.602318	-.1853877
_cons	1.192296	.7205439	1.65	0.098	-.2199444	2.604536
mills						
lambda	-.8664386	.3992843	-2.17	0.030	-1.649021	-.0838558
rho	-0.92910					
sigma	.93255922					

Podemos comprobar que el retorno de la educación usando el comando `heckman` de Stata es el mismo obtenido que cuando aplicamos la metodología de Heckman paso a paso.

#### 4) Estime de manera conjunta la ecuación de selección y la ecuación de interés del modelo de Heckman usando Máxima Verosimilitud. Comente.

En la mayoría de casos, es preferible estimar el modelo completo (tanto la ecuación de selección como la ecuación de interés) de manera conjunta a través de Máxima Verosimilitud.

```

.
. heckman lwage educ exper, select (dwork= age educ kids mtr)

Iteration 0:  log likelihood = -922.95945
Iteration 1:  log likelihood = -914.27456
Iteration 2:  log likelihood = -913.56337
Iteration 3:  log likelihood = -913.56101
Iteration 4:  log likelihood = -913.56101

Heckman selection model                                Number of obs    =       753
(regression model with sample selection)               Censored obs     =       325
                                                        Uncensored obs   =       428

Log likelihood = -913.561                             Wald chi2(2)     =       22.50
                                                        Prob > chi2      =       0.0000

```

	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
lwage						
educ	.0658159	.0166346	3.96	0.000	.0332126	.0984192
exper	.0117675	.0040935	2.87	0.004	.0037444	.0197906
_cons	.6685864	.2350055	2.84	0.004	.2079841	1.129189
dwork						
age	-.0132621	.005939	-2.23	0.026	-.0249022	-.0016219
educ	.0639306	.0217446	2.94	0.003	.0213119	.1065492
kids	-.1525918	.0995874	-1.53	0.125	-.3477796	.042596
mtr	-2.291885	.5375647	-4.26	0.000	-3.345493	-1.238278
_cons	1.595958	.6237306	2.56	0.011	.3734682	2.818447
/athrho	-1.219374	.1181811	-10.32	0.000	-1.451005	-.9877435
/lnsigma	-.1631751	.0500129	-3.26	0.001	-.2611986	-.0651515
rho	-.8394695	.0348978			-.8958914	-.7563985
sigma	.8494425	.0424831			.770128	.9369255
lambda	-.713081	.0605756			-.8318071	-.594355

```

LR test of indep. eqns. (rho = 0):   chi2(1) =    28.64   Prob > chi2 = 0.0000

```

Una ventaja es la estimación por MV es que obtenemos de manera directa el resultado del test LR para comprobar si los errores de la ecuación de selección y los errores de la ecuación de interés están correlacionados.

De los resultados, debemos notar que se rechaza la probabilidad de que  $\rho$  sea igual a cero, este  $\rho$  mide el grado en que el sesgo de selección es importante, si es igual a cero significa que no hay sesgo de selección. Por tanto, comprobamos que sí existe sesgo de selección en la estimación por MCO.

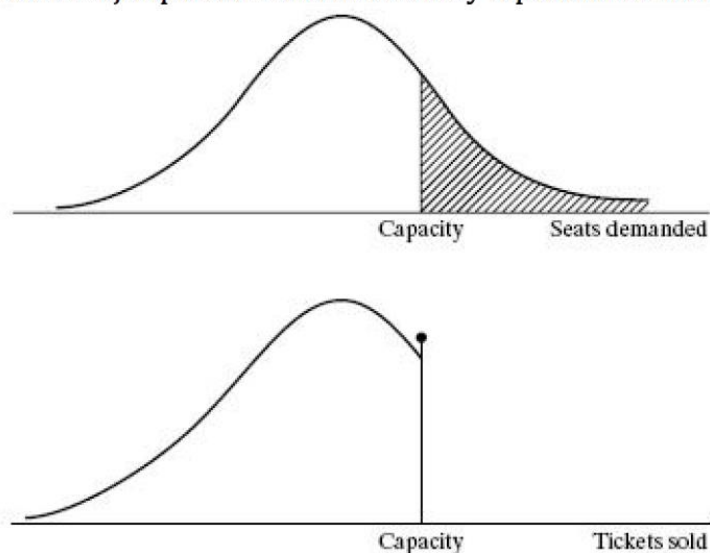
### Pregunta 3 (Opcional) – Tobit<sup>3</sup>

Analizaremos el caso de las variables censuradas. En esta oportunidad analizaremos un caso muy usual que es el de gasto. Este tipo de variables, cuando no son bienes totalmente indispensables (agua o electricidad, por ejemplo), por lo general se truncan en el valor cero porque no se consume en el hogar (es decir, ese mes o año que se consultó simplemente nadie consumió dicho bien).

Frente a un problema de este tipo se pueden hacer dos cosas. La primera es trabajar bajo supuestos. El primero es “asumir” que el gasto es cero (es decir imputar el valor cero) o, en segundo lugar, trabajar únicamente con los datos disponibles (con lo cual nuestros resultados son sólo válidos para el grupo que consume).

Una alternativa es usar un modelo Tobit el cual asume una distribución normal para los datos faltantes. Tal como se observa en el gráfico 1. En este caso reemplazamos un supuesto por otro. Suponemos que en caso de no existir el “límite” la distribución se habría comportado como una variable normal.

Gráfico 1. Ejemplo con venta de boletos y capacidad de auditorio



En este ejercicio compararemos los resultados de un modelos de regresión haciendo estas dos formas de lidiar con los datos faltantes y después usando un modelo Tobit.

El objetivo de este ejercicio es explorar los determinantes del consumo de tabaco<sup>4</sup>. Para ello, la base de datos a utilizar es tobacco.dta, que contiene información sobre 2724 hogares de una muestra de gastos recopilada por el Instituto Nacional de Estadística de Bélgica para 1995/1996. A continuación se describen las variables de la base:

- bluecol : dummy, 1 si el jefe es trabajador "blue collar"
- whitecol : dummy, 1 si el jefe es trabajador "white collar"
- flanders : dummy, 1 si vive en Flanders
- walloon : dummy, 1 si vive en Wallonie

<sup>3</sup> El desarrollo completo de esta pregunta se encuentra en el do-file "tobit\_pregunta3\_opcional".

<sup>4</sup> Este ejercicio se basa en Verbeek (2000), A Guide to Modern Econometrics. New York: Wiley Press.

- nkids : número de niños de más de 2 años de edad
- nkids2 : número de niños de menos de 3 años de edad
- nadults : número de adultos en el hogar
- ln<sub>x</sub> : log del gasto total
- share2 : participación del gasto en tabaco en el total
- share1 : participación del gasto en alcohol en el total
- nadlnx : número de adultos x log del gasto del total
- agelnx : edad x log del gasto total
- age : edad del jefe (0-4): 0 = 0-30 años; 1=30-40; 2=40-50; 3=50-60; 4=>60
- ln<sub>x</sub><sup>2</sup> : ln<sub>x</sub> al cuadrado - age<sup>2</sup> : edad al cuadrado

Estimaremos una curva de Engel para el gasto en tabaco, de acuerdo a la siguiente especificación:

$$w = x'\beta + u$$

Donde  $w$  es share2 y  $x$  es un vector de variables explicativas que incluye age, nadults, nkids, nkids2, ln<sub>x</sub>, agelnx y nadlnx. Se pide realizar lo siguiente:

- a) Analice la variable share2. ¿Qué características presenta?
- b) Estime el modelo (1) por MCO. ¿Qué problema potencial enfrentan los coeficientes estimados?
- c) Estime el modelo reemplazando los ceros con missing. ¿Qué problema potencial enfrentan los coeficientes estimados?
- d) Ahora utilice un Tobit para estimar el mismo modelo. Compare los resultados con los de MCO.
- e) Utilizando el modelo Tobit estimado, halle y compare los valores ajustados de la variable dependiente para los siguientes casos:
  - $E[y^*/x]$  : valor esperado de la variable latente
  - $E[y/x, y > 0]$  : valor esperado de la variable truncada
  - $E[y/x]$  : valor esperado de la variable censurada
  - $Pr[y^* > 0]$  : probabilidad de no censura en  $y^*$
- f) Estime los siguientes efectos marginales y compárelos con los de MCO:
  - a.  $\partial E[y^*/x] / \partial x_k$ : Efecto sobre la variable latente
  - b.  $\partial E[y/x, y > 0] / \partial x_k$ : Efecto sobre el valor esperado de la variable truncada
  - c.  $\partial E[y/x] / \partial x_k$  : Efecto sobre el valor esperado de la variable censurada
  - d.  $\partial Pr[y^* > 0] / \partial x_k$  : Efecto sobre la probabilidad de no censura de  $y^*$